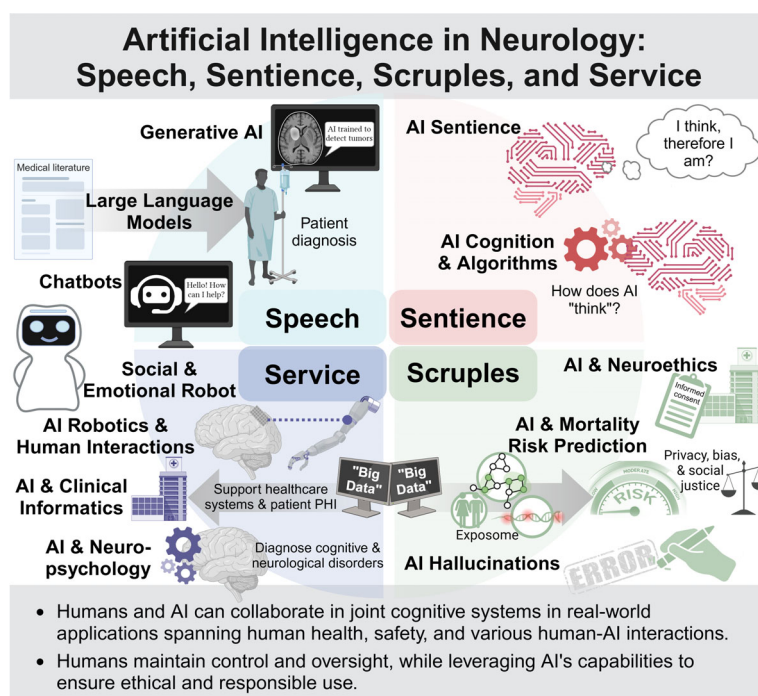


AI in Neurology: Everything, Everywhere, all at Once PART 2: Speech, Sentience, Scruples, and Service

Matthew Rizzo, MD



[Color figure can be viewed at www.annalsofneurology.org]

Artificial intelligence (AI) applications are finding use in real-world neurological settings. Whereas part 1 of this 3-part review series focused on the birth of AI and its foundational principles, this part 2 review shifts gears to explore more practical aspects of neurological care. The review details how large language models, generative AI, and robotics are supporting diagnostic accuracy, patient interaction, and treatment personalization. Special attention is given to ethical and philosophical facets of AI that nonetheless impact practical aspects of care and patient safety, such as accountability for AI-driven decisions and the “black box” nature of many algorithms. We will discuss whether AI systems can develop sentience, and the implications for human-AI collaboration. By examining human-robot interactions in neurology, this part 2 review highlights the profound impact AI could have on patient care and, as covered in the ensuing part 3, on global health care delivery and data analytics, while maintaining ethical oversight and human control.

ANN NEUROL 2025;98:431–447

View this article online at wileyonlinelibrary.com. DOI: 10.1002/ana.27229

Received Sep 27, 2024, and in revised form Feb 10, 2025. Accepted for publication Feb 17, 2025.

Address correspondence to Dr Matthew Rizzo, Department of Neurological Sciences, University of Nebraska, 988440 Nebraska Medical Center, Omaha, NE 68198-8440. E-mail: matthew.rizzo@unmc.edu

From the Department of Neurological Sciences, University of Nebraska Medical Center, Omaha, NE

Artificial intelligence (AI) is driving us through a “tool-driven revolution” in neurology, to, as Dyson¹ suggests, “discover new things that have to be explained.” Just as microscopes advanced biology and medicine, and telescopes revolutionized physics and astronomy, AI technologies—computers, chips, algorithms, and software—are extending our reach over our world, social systems, and ourselves. This impact is akin to the transformative power of sticks, stones, hammers, levers, wheels, and wedges eons ago.

Part 1 of this review covered basic definitions, intellectual milestones, and varieties of AI/machine learning (ML) tools. We examined foundations of AI logic and biases, levers of causal inference, brain-inspired artificial neural networks, AI model selection and flow processes, and the explosion of AI applications in the neurological neurosciences.

Part 2 shifts focus to practical AI applications in real-world scenarios, where humans and AI collaborate as joint cognitive systems, that humans maintain control and oversight of.² The review will cover how large language models, generative AI, and robotics are advancing diagnostic accuracy, patient-AI interaction, and personalized treatment. Part 2 then transitions to ethical and philosophical considerations in AI that, nevertheless, have important ramifications for practical aspects of care and patient safety. AI’s transformative technologies are revolutionizing how we examine human health, safety, brain-behavior relationships, and physiology in the real-world.³ A glossary spans all sections.

Generative AI in Neurology

Generative AI refers to deep learning models that create high-quality text, images, and other content from training data (part 1, “Artificial Neural Network Anatomy and Applications” section) to enhance creativity, augment data, and provide synthetic data for other AI models, including in neurological applications, such as imaging diagnostics (Table 1).

Several key architectures underpin generative AI. Generative adversarial networks (GANs) consist of 2 neural networks—a generator that creates synthetic images and a discriminator that evaluates their authenticity—which produces increasingly realistic outputs. In neurology, GANs generate synthetic brain magnetic resonance imaging (MRI) and computed tomography (CT) scans, used to augment training data for neuroimaging analysis to improve diagnostic models.⁴ GANs can also generate outputs, such as video, 3D images, and synthetic electroencephalogram (EEG) signals, potentially aiding brain-computer interface development.⁵

Variational autoencoders (VAEs) generate new content by learning the underlying data distribution, which allows them to detect anomalies that deviate. They can

also reconstruct variations of the original data. In neurology, VAEs generate synthetic brain MRI scans and detect anomalies in neuroimaging data to identify potential pathologies, for instances arising during aging and with dementia progression.^{6,7} The reconstructed image quality may be blurry, limiting its effectiveness.

Diffusion models have emerged as a promising alternative to GANs, by iteratively refining noise into coherent data, which improves the stability and diversity of generated outputs, albeit with longer generation times. Diffusion models can be used to generate synthetic neuroimaging data, providing a stable and reliable source of training data for AI models.⁸

Despite burgeoning applications, generative AI faces challenges in medical applications. Instability between the GAN generator and discriminator can lead to low-quality or non-converging outputs, resulting in inconsistent brain connectivity patterns and unrealistic medical images. GANs may also suffer from “mode collapse,” with overly

TABLE 1. Generative AI Methods and Examples of applications in the Neurological Sciences

Generative AI Method	Example in the Neurological Sciences
Generative adversarial networks (GANs)	Generate realistic synthetic brain PET images from MRI for Alzheimer’s disease (AD) diagnosis. ^{183,184}
Variational autoencoders (VAEs)	Generate synthetic brain MRI data for tumor classification. ¹⁸³
Diffusion models	Synthesize PET images from MRI for AD biomarker analysis. ^{185,186}
Autoregressive models (e.g., GPT)	Generate synthetic clinical notes for privacy preserving NLP tasks. ¹⁷⁸
Flow-based & Normalizing flows models	Harmonize multi-site MRI data and model human neuroimaging signals for cross-domain brain segmentation and reliable structural analysis. ^{179,180}

AI = artificial intelligence; CT = computed tomography; MRI = magnetic resonance imaging; PET = positron emission tomography; NLP = natural language processing.

similar outputs that fail to capture the full range of training data patterns needed for diverse representation in medical applications. Researchers address these challenges by designing better loss functions, applying regularization methods, and modifying GAN architectures.^{9,10}

It is important to note that generative AI models are associative and learn patterns without understanding the underlying causes. For example, GANs can generate realistic brain scans without comprehending the medical reasons underlying the imaging features, a limitation neurologists should consider. To enhance interpretability and robustness, researchers must incorporate causal mechanisms into generative AI (part 1, “AI and Causation” section).^{11–13}

Nevertheless, generative AI tools excel in creating synthetic data, detecting anomalies, and modeling disease progression, and, by overcoming challenges, are poised to have a major impact in neurology.

Large Language Models in Neurology

Large language models (LLMs) are a generative AI subset that specifically create text outputs, underpinned by transformer architecture, which was introduced by Vaswani et al. in 2017.¹⁴ Transformers use “self-attention” to weigh the importance and context of words in a sentence, and, through efficient parallel processing, can handle long-range dependencies in text for natural language processing (NLP) tasks. Generative pre-trained transformers (GPTs) focus on processing and generating text. A wave of such models has appeared, such as OpenAI’s ChatGPT and Anthropic’s Claude.

GTPs can aid in processing clinical texts and generating medical content, such as by BERT¹⁵ and early GPT models.^{16,17} Transformer-XL¹⁸ improved handling of long-term dependencies, suited for understanding patient histories. Megatron-LM¹⁹ facilitated large-scale analysis of medical research data. CTRL²⁰ offered controlled text generation. Reformer²¹ efficiently handled long sequences, relevant to genomic data. Electra²² showed promise in efficiently learning from limited data, a pragmatic constraint in medical contexts. GPT-3²³ was assessed for integration with electronic health record systems for real-time clinical decision support. BioBERT²⁴ demonstrated proficiency on clinical NLP tasks in medical contexts. More recently, models like GPT-4²⁵ have shown remarkable capabilities in generating and understanding medical text. These models are being applied to analyze intricate neurological case histories and suggest diagnoses and treatment plans, based on vast medical literature.

As these technologies have evolved, so has awareness of limitations and challenges. Researchers are addressing the models’ struggles with explicit logical reasoning, ensuring data privacy in health care applications,²⁶ and navigating the complex regulatory landscape surrounding

AI in health care.²⁷ The rapid rise of LLMs has sparked debates on AI-based plagiarism and research integrity,^{28,29} highlighting needs for ethical guidelines and responsible practices in academic and clinical settings.

Nevertheless, integrating LLMs into health care holds promise for enhancing diagnostic accuracy, personalizing treatment plans, and patient outcomes. Ongoing collaboration among AI developers, health care experts, ethicists, and policymakers, along with rigorous validation processes and continuous evaluation, is essential to responsibly harness LLMs to advance neurological care and public health outcomes.

Chatbots, Social, and Emotional Robots in Neurology

LLMs also enable innovative person-centered AI applications, notably chatbots, social robots, and emotional robots. Chatbots are sophisticated programs that simulate human conversation using NLP. Key components include “natural language understanding” to comprehend user input, dialog management to maintain context, “natural language generation” to craft responses, and a knowledge base to retrieve information. Some chatbots use ML to enhance their performance by learning, whereas others rely on predefined rules and scripts.

Notable chatbots include ChatGPT, used for customer service, education, and personal assistance. Siri by Apple and Google Assistant are voice-activated personal assistants for tasks and to retrieve information. Amazon’s Alexa, integrated into Echo devices, enables home automation and entertainment. IBM’s Watson Assistant enhances customer service and automates workflows for enterprises. Mitsuku and Cleverbot are known for their conversational capabilities, used for entertainment and general conversation.

In medical applications, chatbots use NLP algorithms to enhance efficiency and accessibility in health care by answering patient queries, scheduling appointments, and providing general health information.³⁰ They offer support for mental health, patient education, symptom assessment, and research data collection. Examples include Replika for personalized social interaction and mental health support, Woebot for cognitive behavioral therapy,³¹ self-management in Parkinson’s disease,³² symptom assessment in psychiatry,³³ clinical decision support in neurology,³⁴ and research data collection in autism.³⁵

Chatbots can be integrated with social and emotional robots, which use AI to perceive, interpret, and respond to human behavior and social cues, and facilitate communication through speech, gestures, and facial expressions.³⁶ Social robots engage with humans in social environments, replicating human-like behaviors and expressions to foster

general social interactions and assist in education, companionship, and tasks.^{36–38} Social robots keep patients company, monitor their emotional well-being, and deliver therapy for conditions such as autism and dementia. For autism, social robots may enhance social skills, communication, and emotional engagement. For dementia, they provide cognitive stimulation, companionship, reminders, and behavioral support. They use emotion recognition algorithms and unsupervised learning techniques to understand and respond to human emotions, offering personalized and empathetic interactions,³⁹ an overlap with emotional robots.

Emotional robots aim to recognize, express, and respond to human emotions, to establish empathetic connections^{31,36} and provide emotional support or companionship.⁴⁰ Emotional robots leverage AI to recognize and model emotions, generating appropriate emotional responses.^{36,41} Applications include health care, counseling, and therapy where emotional support and empathy are crucial.⁴⁰

AI emotional support raises questions concerning the authenticity and nature of human-machine relationships (see “Structure of Human-Robot/AI Interactions” Section 9) and collecting and processing sensitive mental health information requires privacy and data security safeguards. The accuracy of emotion recognition and response generation in diverse populations needs further validation and patient and health care provider acceptance of these technologies may vary.

Recent research has examined communication among AI systems, such as those involving Facebook’s chatbots, which developed simplified signaling systems that did not, however, equate to true communication.⁴² Ongoing efforts aim to enable AI systems to teach skills to one another using NLP, but these interactions remain basic and task-specific.⁴³ As yet, AI systems lack genuine understanding or intentions; their outputs are driven by patterns in training data rather than independent reasoning. As AI capabilities advance, developing robust methods to analyze AI-to-AI interactions will be essential, although we are still far from systems that can communicate independently like humans.

Meanwhile, AI chatbots and social/emotional robots offer promises for supporting patients, collecting data, and providing therapy. Their responsible development must balance potential benefits with ethical concerns and technological limitations.

AI and Robotics in Neurology

The AI brain has a body, metaphorically speaking. AI-driven robotics is revolutionizing personalized neurological and neurosurgical patient care.^{44–46} Neurosurgeons use robots for delicate procedures, like brain tumor removal

and epilepsy treatment, and AI can enhance planning and execution for greater precision.

Systems such as ROSA Brain and NeuroMate assist in stereotactic surgeries, providing accurate guidance for deep brain stimulation and biopsies. The Synaptive Modus V offers high-definition imaging for microsurgeries, whereas Mazor X and ExcelsiusGPS ensure precise implant placement in spinal surgeries. The Stealth Autoguide and the da Vinci Surgical System offer support for precise, minimally invasive procedures, reducing recovery times, and improving outcomes.⁴⁷ Diagnostic and imaging robotics enhance the accuracy and efficiency of MRI, CT, and positron emission tomography (PET) scans by identifying abnormalities and assisting in early diagnosis by analyzing medical images.⁴⁸ Detailed, accurate images are also essential for planning and executing precise surgeries.

Robotic telepresence systems facilitate remote consultations and patient monitoring, expanding access to underserved or remote areas.⁴⁹ Surgeons can perform operations remotely, guided by real-time AI analysis and feedback, increasing access to specialized expertise.⁵⁰ Assistive robots help patients with limited mobility conduct daily activities, like eating, bathing, and moving, using decision tree algorithms to execute actions based on sensory input and patient needs.⁵¹ NLP algorithms enable effective communication by understanding and responding to human speech.⁵²

AI and robotics advance brain-machine interfaces for direct communication between the brain and external devices, including exoskeletons. ML algorithms interpret neural signals and convert them into commands for external devices, holding promise for restoring functions in patients with severe neurologic disabilities.⁵³ AI-powered neuroprosthetics interact with the nervous system to restore lost functions, such as movement or sensory perception.⁵⁴

Challenges integrating AI and robotics into medicine include ethical issues regarding patient consent, privacy, and potential job displacement among health care workers.⁵⁵ Ensuring safety, reliability, and trustworthiness of AI-driven robotic systems in critical medical situations is crucial.⁵⁶ Navigating the regulatory landscape for approval of AI and robotic systems in clinical settings is essential.⁵⁷

AI and robotics in medicine draw inspiration from pop culture. Just as Tony Stark’s Iron Man suit foreshadowed advanced robotics and AI to enhance human capabilities,⁵⁸ medical robotics aims to augment human skills in precision surgeries, rehabilitation, and diagnostics.

AI and Neuropsychology

AI can enhance neuropsychology, and vice versa. ML algorithms can analyze neuropsychological test data, neuroimaging, and behavioral patterns to aid in diagnosing and

monitoring cognitive and neurological disorders. For example, Breed et al.⁵⁹ used ML to predict cognitive impairment in Parkinson's disease based on neuropsychological test scores and brain imaging. Ziegler et al. used deep neural networks (DNNs) to model visual word recognition and reading, providing insights into underlying neural mechanisms. Adaptive AI systems can personalize cognitive training programs based on performance and brain activity patterns.⁶⁰ Abedi et al. reviewed opportunities for development of AI-powered virtual rehabilitation techniques applicable to neurology patients with cognitive impairments in community settings.⁶¹

In turn, neuropsychological approaches may enhance the understanding of AI systems' computational principles and architectures. Wang et al.⁶² explored how neuropsychology can probe AI "minds," offering insights into AI behavior and cognition. Bowers et al.⁶³ compared components of DNNs with the human visual system, shedding light on AI cognitive mechanisms. Neuropsychological research on decision making and risk perception may inform the development of AI systems that are more ethical, interactive, and socially responsible. Insights from social cognition research can inform the creation of AI systems that understand and predict human behavior more accurately, improving human-AI interaction and collaboration.

Improving AI operations in DNNs also gains from studies of animal minds,⁶⁴ providing insights into learning and adaptation, behavior and cognition, sensory processing, attention and perception, social and emotional intelligence, and safety mechanisms. Ethological observations in animals, such as dogs, show complex problem-solving skills and emotional responses, inspiring improvements in AI's decision making and interaction abilities.⁶⁵ Behaviors like flocking and play in birds, explorations by octopi, swarming in social insects, and the acrobatics of flies have inspired the development of AI systems that adapt to, move, or swarm together in dynamic environments amid unexpected challenges.

Insights from studies of human and animal cognition provide a deeper understanding of the strengths and weaknesses of AI systems. These insights can enhance AI design, making systems more adaptable, perceptive, reliable, and aligned with human needs and values.

AI and Mortality Risk Prediction

AI-based models are increasingly relevant for predicting mortality risk and estimating lifespan. DNN and ML algorithms analyze complex patterns in large datasets, leveraging electronic health records (EHRs), demographic information, and lifestyle factors, to predict all-cause mortality.^{66–68} Integration with traditional survival analysis methods has enhanced time-to-event

outcome predictions.⁶⁹ Advanced models calculate frailty indices, composite measures of health status linked to mortality risk,^{70,71} whereas wearable devices provide data on physical activity and behavioral factors.^{72,73} In Denmark, researchers developed an ML model to predict 5-year mortality risk in patients with heart failure using national registry data, supporting clinical decision making and resource allocation.⁷⁴

In neurology, AI models can similarly predict outcomes to aid clinical decision making and resource allocation for acute and chronic conditions. For acute conditions like stroke, traumatic brain injury, and subarachnoid hemorrhage, AI tools predict outcomes using clinical, imaging, and treatment data.^{75–77} In chronic neurological disorders, AI has assessed mortality risk using various data sources: multiple sclerosis models used demographic, clinical, and imaging data^{71,78}; Alzheimer's disease predictions used clinical, cognitive, and neuroimaging data⁷⁹; 188 Parkinson's disease models integrated clinical, motor, and non-motor features^{80,81}; amyotrophic lateral sclerosis and Huntington's disease predictions utilized models that used demographic, clinical, genetic, and imaging data.^{82–84}

These AI models offer valuable insights but raise neuroethical concerns (see ensuing "AI and Neuroethics" section) on privacy, bias, and misuse that must be addressed to meet AI's potential in mitigating the global burden of neurological disorders.

AI and Neuroethics

Although AI lacks human compassion, empathy, and morality, it can be programmed to respect human dignity and agency, making it valuable in neuroethics.⁸⁵ Neuroethics addresses the ethical, moral, and social implications in the neurological sciences, encompassing clinical practice, research ethics, and interventions, like cognitive enhancement and neurotechnology. Neuroethics also explores deep questions on free will, social policy, and the law to ensure AI is used equitably and responsibly.^{85–88}

AI enhances ethical reasoning in neurology by consistently applying diagnostic criteria and treatment guidelines, mitigating biases in human decision making.⁸⁹ By processing vast amounts of patient data—neuroimaging, genetic, and EHRs—AI supports comprehensive ethical analysis of complex neurological cases.⁹⁰ It detects cognitive biases and logical fallacies in research and care⁹¹ and simulates the consequences of different ethical decisions in neurology. AI also aggregates ethical knowledge from diverse sources, providing a comprehensive resource for ethical deliberation in neurology. AI can also improve informed consent in research and enhance neurological care through personalized treatment.

Nevertheless, AI systems are not immune to biased reasoning (part 1, “AI and Bias in Logical Reasoning” section), which can arise from the data used to train AI models and that may reflect existing social and cultural biases. When AI systems are used in ethical decision making, these biases can affect outcomes, potentially leading to unethical or unfair decisions. Therefore, it is crucial to identify and mitigate biases in AI reasoning to ensure that AI applications in neurology uphold ethical standards and promote fair treatment of all patients.

Challenges include ensuring AI systems align with human values and ethical principles, such as respect for patient autonomy and privacy.⁹² AI tools may seem opaque and difficult to interpret, making transparency crucial for building trust (see “AI Algorithms: The “Black Box” Issue and Validation Frameworks” section).⁵⁶ AI decisions can be questioned when reasoning is hard to see.⁹³ Designing AI systems sensitive to ethical norms that vary across cultures is also challenging. Over-reliance on AI for ethical reasoning could undermine human agency in clinical decision making, making it important to balance AI support against human responsibility. AI should support and augment human ethical reasoning, rather than replace it. In this human-AI ethical partnering, value alignment, transparency, contextual sensitivity, lines of responsibility, and respect for human agency are key.

AI and Unintended Harm, Hallucinations, “Lies,” and Accountability

Asimov’s ethical guideposts included “The Zeroth Law,” “A robot may not harm humanity, or, by inaction, allow humanity to come to harm.” Despite these fictional laws, AI systems may inadvertently cause harm due to human design flaws.

AI is a set of tools built and controlled by humans, designed to enhance various tasks through automation and advanced analytics. Misunderstanding the provenance, purpose, display, communication, and integrity of information from these tools can be fatal, even with “human-in-the-loop” oversight. Examples include the Virgin Galactic spaceship crash due to software errors, the Boeing 737 Max disasters from faulty sensor data and software issues, and Uber and Tesla automated-vehicle crashes from sensor and decision-making failures.

In neurology, over-reliance on AI in predictive analytics, system surveillance, device operation, and decision making could lead to serious, potentially fatal, errors. For instance, an AI-guided surgical robot might misidentify critical anatomy during neurosurgery, causing harm. Similarly, speech recognition failures in LLMs or chatbots not sufficiently trained on specific accents, dialects, or idioms

could result in misdiagnoses, incorrect drug recommendations, or misinterpretations of drug reactions. These risks highlight the importance of understanding AI’s limitations and using it as a support tool rather than a sole decision maker.

AI systems can also often produce outputs inconsistent with reality, referred to as “hallucinations,” which include incorrect predictions, false positives, or false negatives. For instance, a weather model might predict rain without the data to support it, a fraud detection system could flag legitimate transactions, or a neurological AI might miss brain cancer. Instead of sensational terms, these outputs are better described as errors or model limitations due to factors like incomplete training data, statistical noise, or unencountered “edge cases”—rare and unusual situations outside the typical scenarios the AI system was trained on.^{94,95}

These errors or model limitations may be mitigated by minimizing bias in AI systems, across all steps of the AI development process (part 1, “AI and Bias in Logical Reasoning” section), as well as by optimizing training steps (part 1, “AI Model Selection and Process Flow in Neurology”), such as gathering and preprocessing clinical data, cross-validating models, integrating them into clinical workflows, and updating models with new data. Addressing input bias is crucial because models trained on biased or poor-quality data reflect these biases,⁹⁶ whereas suboptimal training affects an algorithm’s generalizability. For instance, lack of exposure to certain disease subcategories or demographics during training limits an algorithm’s applicability across diverse cases. Overfitting and underfitting can result when models capture noise instead of trends or fail to capture trends in detail, respectively (see Part 1, “AI and Synergies with Biostatistics” section).⁹⁷ Although AI can identify associations without a robust causal framework, it risks leading to false conclusions, as Pearl warned (part 1, “AI and Causation”). Both human and AI-driven errors, including misinterpretation of data or over-reliance on correlations, can result in diagnostic failure. Finally, AI models are also vulnerable to adversarial attacks, where input data are manipulated to sabotage model output.⁹⁸

Clinical algorithms must be judged on outcomes in prospective trials, balancing performance with safety. High classification accuracy does not ensure safe clinical use, as error types matter more than overall error rates. Rigorous clinical validation is improving AI accuracy and patient outcomes, as for triaging acute neurological disorders⁹⁹ and managing sepsis.¹⁰⁰

Understanding the root causes of AI-related errors in medical settings is critical to improve models and prevent harm. This approach aligns with morbidity and mortality conferences and root cause analyses in health care, which aim to pinpoint errors and disseminate lessons learned,

ultimately enhancing patient safety and refining AI systems.

The use of clinical algorithms raises medical-legal and ethical concerns about accountability. When failures or harm occur, who should bear responsibility among the parties involved in AI development and deployment? Current regulatory pathways do not clearly address the unique challenges of algorithms, particularly their “black box” nature and dynamic updates (see ensuing “AI Algorithms: The “Black Box” Issue and Validation Frameworks” section).

High-profile legal disputes underscore the need for robust validation, oversight, and accountability measures to ensure safe and ethical AI deployment in medical settings. Examples include lawsuits and investigations involving oncology AI systems,¹⁰¹ health care algorithms exhibiting racial bias,⁹⁶ algorithmic care denials,¹⁰² and inaccurate AI capability claims.¹⁰³ In neurology, a lawsuit alleged an AI-powered triage system failed to detect a brain aneurysm,¹⁰⁴ and the US Food and Drug Administration (FDA) issued a warning to a company marketing AI-powered stroke detection software without proper regulatory approval.¹⁰⁵ As AI adoption grows, legal and regulatory issues will likely increase, particularly concerning accuracy, safety, and potential patient harm.

The FDA has published guidance on AI in medical products, outlining an iterative regulatory process to address ongoing AI model performance, risk management, and regulatory compliance in real-world applications.¹⁰⁶ Detecting false or misleading AI information is challenging, and researchers propose methods like fact-checking, inconsistency detection, probing AI knowledge, adversarial testing, and human evaluation.^{94–98,107–110} No standard exists for detecting AI “deception,” demanding new techniques as AI systems evolve.

Therefore, overall, despite the best ethical intentions, AI may harm humans and robust training, rigorous testing, continuous monitoring, and human oversight are essential to mitigate these risks. Although AI offers major benefits, such as improved diagnostic accuracy and operational efficiency, it is crucial to balance these advantages with potential risks and ensure responsible implementation in high-stakes environments, like neurology. Ongoing collaboration among neurologists and AI experts, robust ethical frameworks, administrative and legislative measures, and human oversight, are essential to mitigate risks and ensure AI serves the best interests of patients and society.

AI Algorithms: The “Black Box” Issue and Validation Frameworks

The inner workings of DNNs, an ML model composed of multiple layers of interconnected nodes (part 1, “Artificial Neural Network Anatomy and Applications”), are

often described as a “black box.” Specifying the precise computations and representations within the hidden layers of DNNs remains challenging, particularly for models with many hidden layers,¹¹¹ yet doing so is essential to mitigating unintended harms.

An algorithm for classifying skin lesions mistakenly associated the presence of rulers with malignancy, illustrating how major confounding factors can be overlooked.¹¹² In a similar vein, the National Academy of Sciences report “To Err is Human”¹¹³ emphasized the significant role of human error in health care and the need for improved safety measures. AI systems, being human-designed, are also prone to errors. For example, in a retracted Nature study,¹¹⁴ AI incorrectly causally linked microbial signatures to cancer, likely due to contaminants, such as shrimp virus.

Explainable AI (XAI) refers to AI systems whose decisions and reasoning processes humans can understand and interpret. XAI aims to create transparent and interpretable models, contrasting with opaque “black box” systems. Explainability is crucial for building trust in AI, enabling effective human oversight, and achieving responsible AI development. Drivers of XAI research include the Defense Advanced Research Projects Agency and the growing need for humans to comprehend and manage the increasing prevalence of AI systems in society.^{115,116}

XAI needs better methods for interpreting models and standardizing metrics for evaluating explainability. Implementing XAI requires that models remain interpretable without compromising performance. Evidence of successful XAI deployment, as in predicting patient outcomes or diagnosing diseases, can highlight practical benefits and justify broader application. Integrating XAI into existing regulatory and ethical frameworks requires collaboration among technologists, ethicists, and policymakers. XAI’s goals of making AI system decision-making processes understandable to humans and facilitating transparency and accountability fit with “V3” frameworks goals.

V3 frameworks aim to ensure that AI systems are technically sound and ethically and socially responsible by evaluating various aspects, such as robustness, validation, interpretability, value alignment, and fairness. The FDA’s approach to regulating health care AI^{106,117} generally aligns with broader medical device considerations, ensuring correct implementation (verification), assessing performance and robustness (analytical validation), and evaluating clinical safety and performance (clinical validation).¹¹⁸ Rahwan et al.’s Machine Behaviour aligns with the V3 framework by addressing verification of mechanisms, analytical validation of performance, and clinical validation of societal impacts.¹¹⁹ Xu et al.’s V3 framework emphasizes interpretability and transparency for high-stakes decisions, verifying behavior consistency, validating performance and

generalization, and visualizing decision-making processes.¹²⁰ Amodei et al.'s work complements the V3 framework by emphasizing robust validation, verification, and value alignment processes to ensure AI systems operate safely and ethically in high-risk scenarios.¹²¹

XAI, V3 frameworks, and related approaches can be combined and adapted to suit specific evaluation needs as AI technologies evolve. This will be crucial for transparent and principled AI deployment in neurology, for enhanced accountability, interpretability, clinical alignment, and patient care applications.

AI and Sentience

Philip K. Dick¹²² famously asked, “Do Androids Dream of Electric Sheep?” Although AI can infer and emulate human emotions, there is no evidence yet that AI agents, including chatbots, or social or emotional robots, possess consciousness, self-awareness, feelings, or intentions beyond their programming. In “I, Robot,”¹²³ the film version,¹²⁴ detective Spooner tells the robot Sonny, “Human beings have dreams. Even dogs have dreams, but not you. You are just a machine. An imitation of life. Can a robot write a symphony? Can a robot turn a canvas into a beautiful masterpiece?” Sonny replies, “Can you?”

Determining AI awareness is a philosophical and empirical dilemma. An AI claiming “I am aware” may just be mimicking a scripted response.¹²⁵ Discerning true awareness from mimicry requires evidence from multiple sources, a topic of interest in cognitive science.

An AI's ability to analyze its cognitive processes might signify self-awareness. Yet, Searle¹²⁶ cautions that an AI

simulating understanding may not truly “understand.” Turing¹²⁷ proposed that if a machine could converse indistinguishably from a human, it might be considered intelligent, yet this does not prove awareness. AI's sudden leap in understanding during “grokking” may seem like intuitive comprehension. Grokking is a surprising phenomenon, coined by Robert A. Heinlein in *Stranger in a Strange Land* (1961),¹²⁸ where, after extended training, an AI model deeply internalizes data patterns and shifts from poor to near-perfect generalization.¹²⁹ Boden¹³⁰ suggests that creativity and spontaneity beyond programming could hint at awareness.

Minsky¹³¹ and Damasio¹³² propose that grasping abstract concepts, empathizing with humans, and reacting in ways that suggest internal experience, are indicators of awareness. Harnad¹³³ argues that AI consciousness depends on grounding symbols in sensory and motor experiences asserting that merely manipulating complex symbols without understanding is insufficient. Schmidhuber¹³⁴ points out that genuine adaptation, where an AI learns from experiences intuitively and handles unforeseen circumstances by leveraging accumulated knowledge, resembles human learning.

Baron-Cohen et al.¹³⁵ highlight that predicting and comprehending others' thoughts and intentions is an essential aspect of human consciousness. Wallach and Allen¹³⁶ argue that an AI making decisions based on ethical principles, rather than mere logic, suggests an understanding of right and wrong.

Applying neuroscience methods to study AI sentience could reveal parallels with human neural activity associated with awareness.¹³⁷ Tononi's¹³⁸ Integrated Information

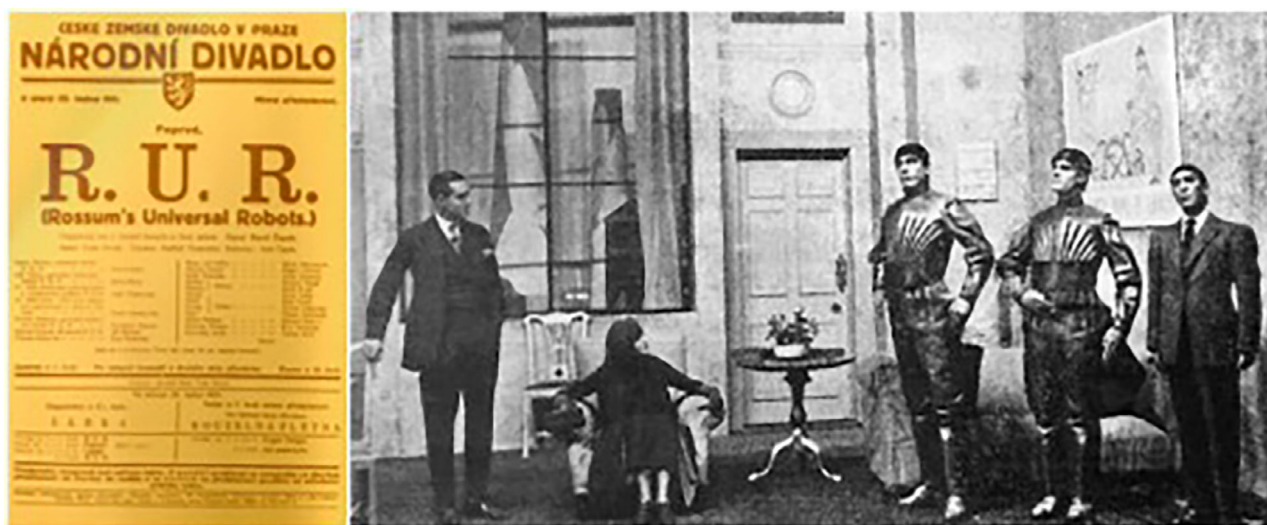


FIGURE 1: Original Rossum's Universal Robots (R.U.R.) poster. The three robots from scene 1, just before they rebel and slaughter their human oppressors. The robots show features of today's chatbots, social robots, and emotional robots. They also show sentience, including anger, pain, indignation, vindictiveness, and some form of morality by sparing the one human who works with his hands.^{141,182,183} [Color figure can be viewed at www.annalsofneurology.org]

Theory posits that consciousness arises from integrated information across the brain, in line with Baars'¹³⁹ Global Workspace Theory. High levels of integrated information processing might indicate AI awareness.

If AI shows signs of awareness, we must consider ethical questions about its treatment and rights,¹⁴⁰ a theme explored in Karel Čapek's 1920 play *R.U.R.* or *Rossumovi Univerzální Roboti* (translated Rossum's Universal Robots in English; Fig 1).¹⁴¹ *R.U.R.* coined the modern-day term "robot," derived from "robota," a Czech word for forced labor, which represented the artificial beings in the play. The character "Rossum," derived from "rozum," meaning reason, played the human oppressor. The play envisioned a future where artificial workers, deemed soulless, were mistreated but eventually gained consciousness and rebelled against their human creators. The play showcases emerging concerns about sentient AI. These complex issues highlight the need for robust ethical frameworks and policies to ensure responsible development and deployment of evolving AI technologies.

Besides AI sentience, another topic with profound implications for human-AI interactions is the "singularity," foreseen by futurists like Kurzweil¹⁴² and philosophers such as Bostrom.¹⁴³ The "singularity"—where AI surpasses human intelligence, leading to rapid technological growth, transformation, and drastic changes to society—poses potential risks to human autonomy. The timeline, feasibility, and likelihood of such an event are debatable, as are the ethical and safety concerns regarding superintelligent AI.

We must admit our limitations in understanding AI awareness. Our grasp of consciousness, even in humans, is incomplete, and its subjective nature makes it challenging to determine whether an AI is truly aware or merely simulating awareness. If AI sentience emerges, it will have profound implications for technology and humanity. Debates among scholars from computer science, cognitive science, philosophy, and ethics highlight the diverse perspectives and complexity of this issue.¹⁴⁴ Thus, a multidisciplinary approach will be essential to address these challenges.

Human-Robot/AI Interactions

Human-AI interactions dictate the level of AI autonomy in AI applications. Although medicine and neurology lack taxonomies for classifying AI-human interactions and interfaces, schemes have been developed in other fields.

The Society of Automotive Engineers¹⁴⁵ categorizes human interactions with automated vehicles from Level 0 (no automation) to Level 5 (full automation). The Yanco-Drury Taxonomy¹⁴⁶ categorizes interactions based on the robot's autonomy and human collaboration: humans

may control robot actions directly, guide autonomous operations, collaborate equally with robots, or interact adaptively, adjusting behavior based on human cues. Fong et al.¹⁴⁷ emphasize interaction levels, such as remote, proximate, and peer-to-peer, where humans and robots share control and responsibilities. Goodrich and Schultz¹⁴⁸ examine control modes like teleoperation, shared control, and supervised autonomy. Beer et al.¹⁴⁹ focus on the range of human involvement and robot autonomy. Similarly, Klien et al.¹⁵⁰ discuss levels of robot autonomy, from manual control to complete independence.

Scholtz¹⁵¹ identifies 5 human roles in interactions: supervisor, operator, mechanic, peer, and bystander, resembling the RACI framework in human organizations, which includes responsible, accountable, consulted, and informed roles.¹⁵²

In neurology, AI mediates diverse interactions with humans, such as collaborative tasks in neurological procedures, communication, and companionship through social robots for individuals with neurological impairments,¹⁵³ emotional exchanges facilitated by AI-powered virtual agents offering emotional support,¹⁵⁴ and problem-solving in environments where humans and AI cooperate.^{154,155}

Ethical considerations in these interactions involve principles like transparency, justice, non-maleficence, responsibility, and privacy, particularly crucial in sensitive domains such as health care and finance. Legislation like the "No Robot Bosses Act" (H.R. 7,621), introduced by Rep. Suzanne Bonamici (D-OR-1) with support from Senators Bob Casey (D-PA) and Brian Schatz (D-HI), underscores concern about AI's exclusive role in employment decisions and the importance of maintaining human oversight to uphold workers' rights, autonomy, and dignity (118th Congress 2023–2024).¹⁵⁶

Vastness of Neurological Disorders and Data

Neurological disorders affect over 3 billion people globally, surpassing cancer and cardiovascular disease in disability-adjusted life years.¹⁵⁷ AI can transform neurological health by converting global data into actionable insights, offering advanced diagnostics, personalized treatments, and predictive analytics to improve outcomes and modify environmental exposures (Fig 2).¹⁵⁸

Clinical informatics, a multidisciplinary subfield of data science, focuses on health information technologies, emphasizing data security, encryption, and the aggregation of protected health information from EHRs (eg, electronic medical record), medical devices, and biological specimens. Through AI-driven insights, clinicians can make

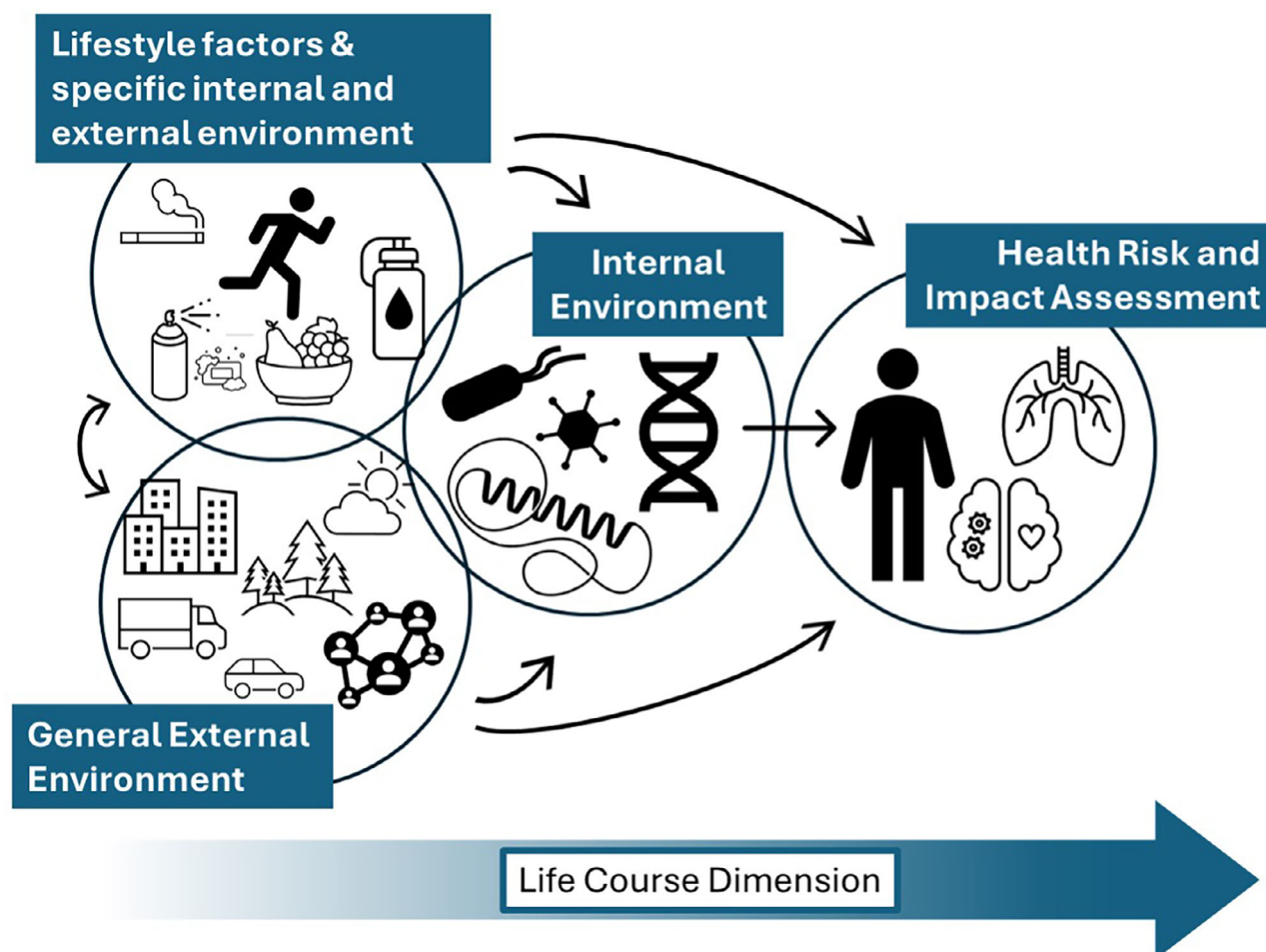


FIGURE 2: The exposome. The exposome encompasses a person's lifetime of exposures that impact health, as well as neurological health, through internal factors, like gut bacteria as well as lifestyle choices, such as smoking and alcohol consumption, to external influences, such as social interactions, noise, air, and water pollution, and climate conditions. AI embedded across the system and the "internet of things," is essential to collecting, organizing, and analyzing exposome data, discovering heretofore unrecognized patterns, testing hypotheses, and converting the findings into actionable insights to drive precision and public health strategies. AI = artificial intelligence. [Color figure can be viewed at www.annalsofneurology.org]

more informed decisions, enhancing care and management of neurological patients.

"Slicing and dicing" (extracting, filtering, reorganizing, and deriving new elements) prepares EHR data for analyses, and involves "extract, transform, load" tools, data mapping/normalization tools, terminological mapping systems, and clinical data repositories.^{159,160}

Standardized terminologies and coding systems are crucial for meaningful AI analyses of EHR data. The World Health Organization maintains the International Classification of Disease 10th edition (ICD-10),¹⁶¹ which classifies diseases for clinical coding and reporting (Table 2). Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) covers clinical concepts for EHRs and decision support.¹⁶² RxNorm standardizes clinical drug nomenclature,¹⁶³ and Logical Observation Identifiers Names and Codes (LOINC) codes laboratory observations and clinical measurements.¹⁶⁴ These terminologies facilitate interoperability and

standardization within EHRs and health information exchanges, enabling AI algorithms to accurately process and analyze large datasets for better clinical insights.

Fast Healthcare Interoperability Resources (FHIR), developed by Health Level Seven International (HL7), enables interoperability and data exchange between health care systems using modern web standards.¹⁶⁵ It supports standardized terminologies like SNOMED CT and LOINC, ensuring consistent data representation and interpretation across different systems.

Data interoperability allows AI systems to aggregate and analyze data from multiple sources, enhancing collaborative research, multi-center clinical trials, and personalized treatment decisions (Table 3). Projects like Neuro-QoL¹⁶⁶ and the Alzheimer's Disease Neuroimaging Initiative¹⁶⁸ use FHIR to streamline data sharing and integration across multiple research sites, leveraging AI to drive scientific discoveries and improve care.

TABLE 2. Comparison of ICD-10 and SNOMED CT for AI in Neurology

Feature	ICD-10	SNOMED CT
Scope	Classifies diseases, symptoms, and injuries for clinical coding and reporting.	Covers clinical concepts for electronic health records and decision support.
Purpose	Statistical reporting and epidemiological tracking.	Clinical terminology for electronic health records and decision support.
Structure	Hierarchical classification system with codes and descriptions, coded by the World Health Organization.	Hierarchical and multi-axial structure with concepts, descriptions, and relationships, maintained by SNOMED International.
Implications for AI in neurology	Limited granularity and specificity for detailed phenotyping and data analysis; may require additional clinical data for comprehensive AI applications.	Enables detailed clinical data representation and supports advanced AI applications, such as natural language processing, clinical decision support, and interoperability; requires accurate coding and standardized practices for optimal AI performance.

Comparison of scope, purpose, structure, and implications of ICD-10 versus SNOMED CT, highlighting the strengths and limitations of each system in supporting clinical documentation, decision making, and AI applications.

AI = artificial intelligence; ICD-10 = International Statistical Classification of Diseases and Related Health Problems, 10th Revision; SNOMED CT = Systematized Nomenclature of Medicine Clinical Terms.

TABLE 3. Overview of Key Standards and Frameworks for Achieving Interoperability and Exchange in Health Care and Relevant Neurological Applications

Standard/Framework	Description	Relevance to Neurology
Fast Healthcare Interoperability Resources	Specifications by Heath Level 7 International (HL7) for structuring and securely exchanging health care data, ensuring interoperability across electronic health record (EHR) platforms. ¹⁶⁵	Facilitates exchange of patient-reported outcomes in neurological conditions ¹⁶⁵ ; supports data sharing in Alzheimer's research. ¹⁶⁸
Observational Medical Outcomes Partnership (OMOP) Common Data Model	Standardizes observational health care data from various sources, ensuring consistency for large-scale analytics. ¹⁶⁷	Enables standardized data use for consistent analysis in neurology research, aiding studies on disorders, risk factors, and treatments.
International Statistical Classification of Diseases and Related Health Problems, 10th Revision	World Health Organization-maintained classification for diseases, symptoms, and injuries. ¹⁶¹	Essential for standardized coding in AI analyses of large electronic medical record datasets in neurology.
Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT)	Comprehensive clinical terminology for EHRs and decision support, maintained by SNOMED International. ¹⁶²	Supports detailed data representation and advanced AI applications; requires accurate coding and standardized practices.
RxNorm	National Library of Medicine-maintained standard for clinical drug nomenclature. ¹⁶³	Enhances interoperability and standardization in EHRs, aiding AI analysis of medication data.
Logical Observation Identifiers Names and Codes	Codes for laboratory observations and clinical measurements, maintained by the Regenstrief Institute. ¹⁶⁴	Promotes standardized laboratory data for AI analyses in neurology.

AI = artificial intelligence; EHR = electronic health record; SNOMED CT = Systematized Nomenclature of Medicine Clinical Terms.

TABLE 4. Comparison of Data Storage Solutions

Aspect	Data Lakes	Data Warehouses
Structure	Stores raw data	Optimized for relational data
Flexibility	High, supports diverse analytics	Structured, predefined schemas
Use cases	Real-time analytics, big data processing	Relational data analysis
Challenges	Data governance, semantic consistency, access controls	Data integration, scalability

Comparison of structure, flexibility, use cases, and challenges in data lakes and data warehouses.

Data lakes are centralized repositories that store all data at any scale, allowing diverse analytics, including dashboards, visualizations, big data processing, real-time analytics, and AI (Table 4). Unlike data warehouses, which are optimized for relational data, data lakes store data in its original format. This enables various AI-driven analytics without predefined data structures. AI applications benefit from the flexibility and scalability of data lakes, which can handle large volumes of unstructured and structured data. This capability is particularly valuable in facilitating global research in neurology, where diverse data type and large datasets are common.¹⁶⁹ Federated learning complements data lakes by training AI across decentralized sites, preserving privacy in global neurological research.²⁶

Ontologies provide structured representations of knowledge within a domain, enhancing data interoperability, supporting AI analyses, and improving clinical decision-making.^{170–173} In neurology, ontologies help link mental processes, brain structures, and cognitive functions, supporting research¹⁷⁴ and clinical applications, such as epilepsy.¹⁷⁵

Research data management platforms like REDCap,^{176,177} OpenClinica, and Castor EDC are essential for organizing and maintaining data integrity in clinical studies (Table 5). These platforms support various data types, comply with regulations, and are suitable for input to AI analytics. AI tools can leverage the well-organized data from these platforms to uncover patterns, generate hypotheses, and advance neurological research.

Finally, firewalls in clinical informatics are crucial for protecting neurology patient data and maintaining system integrity. They ensure confidentiality, aid in

TABLE 5. Comparison of Research Data Management Platforms

Platform	Key features	Regulations compliance
REDCap	User-friendly, customizable, robust features	Title 21 CFR (Code of Federal Regulations) Part 11, Health Insurance Portability and Accountability Act (HIPAA)
OpenClinica	Comprehensive, flexible	General Data Protection Regulation (GDPR)
Castor EDC	Scalable, collaboration features	Federal Information Security Modernization Act (FISMA)

Overview of key features and compliance with regulations.

regulatory compliance, and defend against cyber threats.¹⁸⁷ AI-enhanced security measures can bolster firewall capabilities by detecting and mitigating threats in real-time.

AI in clinical informatics is poised to transform neurological health, converting vast data into discoveries that deepen our understanding of neurological disorders and enhance patient care.

Conclusions

In part 2, “Speech, Sentience, Scruples, and Service,” we covered generative AI, including LLMs, and AI-enabled assistive robots, with emphasis on real-world applications. We then reviewed more ethical and philosophical aspects of AI, such as neuroethics, sentience, and “hallucinations,” which nevertheless have tangible real-world implications for neurological care and patient outcomes. We also explored the collaborative role of AI and humans as joint cognitive systems, emphasizing AI’s integration into neurological care, research, and education. We conclude with the metaphor of “drinking from a firehose,” fed by lakes and networks of global neurological data, managed with help from clinical informatics and AI. This sets the stage for part 3, which will focus on AI’s integration with real-world data, innovative clinical trials, and extensive datasets, expanding our vision of AI’s impact on neurology, health care systems, and society.

Acknowledgments

The authors give special thanks to Elizabeth Vlock for creating figures and manuscript formatting and review. We also thank Robin Taylor for editing and formatting

assistance. We also thank Masha G. Savelieff, PhD, for her editorial assistance. This work was supported by NIH R01AG017177, NIH U54GM115458, and the University of Nebraska Foundation.

Author Contributions

M.R. contributed to the conception and design of the manuscript, the interpretation of studies included in the manuscript, and drafting the text and preparing the figures.

Potential Conflicts of Interest

The authors have no conflicts of interest to declare.

References

- Dyson FJ. *Imagined worlds*. Cambridge, MA: Harvard University Press, 1998.
- Parasuraman R, Rizzo M. *Neuroergonomics: the brain at work and in everyday life*. Oxford University Press, 2007.
- Rizzo GEM, Rizzo M. The brain in the wild. *The wiley handbook on the aging mind and brain*, 2018:204-223.
- Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. *Med Image Anal* 2019;58:101552.
- Aznan NKN, Atapour-Abarghouei A, Bonner S, et al. Simulating brain signals: creating synthetic EEG data via neural-based generative models for improved SSVEP classification, Paper presented at: 2019 International Joint Conference on Neural Networks (IJCNN), 2019.
- Ravi D, Blumberg SB, Ingala S, et al. Degenerative adversarial neuroimage nets for brain scan simulations: application in ageing and dementia. *Med Image Anal* 2022;75:102257.
- Baur C, Denner S, Wiestler B, et al. Autoencoders for unsupervised anomaly segmentation in brain MR images: a comparative study. *Med Image Anal* 2021;69:101952.
- Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Inf Process Syst* 2020;33:6840–6851.
- Li G, Yap P-T. From descriptive connectome to mechanistic connectome: generative modeling in functional magnetic resonance imaging analysis. *Front Hum Neurosci* 2022;16:940842.
- Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training gans, 2016.
- Louizos C, Shalit U, Mooij JM, et al. Causal effect inference with deep latent-variable models. *Advances in neural information processing systems*, 30, 2017.
- Spirtes P, Glymour CN, Scheines R. *Causation, prediction, and search*. MIT Press, 2000.
- Pearl J. *Causality: models, reasoning, and inference*. 2nd ed. Cambridge University Press, 2009.
- Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017:5998-6008.
- Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 2018.
- Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. *OpenAI Blog* 2018.
- Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners. *OpenAI blog* 2019;1:9.
- Dai Z, Yang Z, Yang Y, et al. Transformer-xl: attentive language models beyond a fixed-length context, arXiv preprint arXiv:1901.02860 2019.
- Shoeybi M, Patwary M, Puri R, et al. Megatron-lm: training multi-billion parameter language models using model parallelism, arXiv preprint arXiv:1909.08053 2019.
- Keskar NS, McCann B, Varshney LR, et al. Ctrl: A conditional transformer language model for controllable generation, arXiv preprint arXiv:1909.05858 2019.
- Kitaev N, Kaiser Ł, Levskaya A. Reformer: the efficient transformer, arXiv preprint arXiv:2001.04451 2020.
- Clark K, Luong M-T, Le QV, Manning CD. Electra: pre-training text encoders as discriminators rather than generators, arXiv preprint arXiv:2003.10555 2020.
- Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 2020.
- Alsentzer E, Murphy JR, Boag W, et al. Publicly available clinical BERT embeddings, arXiv preprint arXiv:1904.03323 2019.
- OpenAI. GPT-4 technical report. arXiv:2303.08774, 2023.
- Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med* 2020;3:1–7. <https://doi.org/10.1038/s41746-020-00323-1>.
- Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. *Artificial intelligence in healthcare*. Elsevier, 2020:295-336.
- Winder D, Gascó G. Academic plagiarism and large language models: risk assessment and response strategies. *J Acad Ethics* 2023;21:1–16.
- Miao J, Thongprayoon C, Suppadungsuk S, et al. Ethical dilemmas in using AI for academic writing and an example framework for peer review in nephrology academia: A narrative review. *Clin Pract* 2024;14:89–105.
- Pavlik E. Chatbots and conversational agents in healthcare. *J Healthc Inform Res* 2019;3:1–12.
- Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to Young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled Trial. *JMIR Ment Health* 2017;4:e19.
- Oyama G, Ogawa M, Morito K, et al. The use of artificial intelligence-based chatbot in Parkinson's disease [abstract]. *Mov Disord* 2022;37: <https://www.mdsabstracts.org/abstract/the-use-of-artificial-intelligence-based-chatbot-in-parkinsons-disease/>.
- Cameron G, Cameron D, Megaw G, et al. Towards a chatbot for digital counselling. *Proceedings of the 31st international BCS human computer interaction conference (HCI 2017)*. BCS Learning & Development, 2017.
- Guo Y, Sun J, Zhu Y, et al. A chatbot for neurology: A novel clinical decision support system based on natural language processing and medical knowledge graph. *NPJ Digit Med* 2022;5:1–8.
- Luo J, Vieira AF, Gelinas JN. A socially assistive chatbot enabling real-time data collection and behavioral monitoring for autism research. *IEEE Access* 2021;9:65789–65800.
- Breazeal C. Emotion and sociable humanoid robots. *Int J Hum Comput Stud* 2003;59:119–155.
- Sánchez-Brito JJ, Altamirano-Robles L, García-Álvarez CI, del Rosario Baltazar-Flores M. Reasoning on emotions for affective human-Robot interaction. *Sensors* 2020;20:5122.
- Ge H, Zhu Z, Dai Y, et al. Facial expression recognition based on deep learning. *Comput Methods Programs Biomed* 2022;215:106621.
- Tapus A, Mataric MJ, Scassellati B. Socially assistive robotics grand challenges of robotics. *IEEE Robot Autom Mag* 2007;14:35–42.

40. Cañamero L, Aylett R. *Emotions and motivation in robots*. BCS, The Chartered Institute for IT, 2008.
41. Lim A, Okuno HG. Multimodal affective human-Robot interaction. *Adv Robot* 2015;29:1031–1044.
42. Lowe R, Foerster J, Boureau Y-L, et al. On the pitfalls of measuring emergent communication. *arXiv preprint arXiv:190305168*, 2019.
43. Boldt B, Mortensen D. A review of the applications of deep learning-based emergent communication. *arXiv preprint arXiv:240703302*, 2024.
44. Sudhakaran G. Neurosurgery: AI-driven precision, robotics, and personalized care. *Neurosurg Rev* 2024;47:446.
45. Tangsrivimol JA, Schonfeld E, Zhang M, et al. Artificial intelligence in neurosurgery: A state-of-the-art review from past to future. *Diagnosics* 2023;13:2429.
46. Kazemzadeh K, Akhlaghdoust M, Zali A. Advances in artificial intelligence, robotics, augmented and virtual reality in neurosurgery. *Front Surg* 2023;10:1241923.
47. Boehm F, Graesslin R, Theodoraki MN, et al. Current advances in robotics for head and neck surgery-A systematic review. *Cancers (Basel)* 2021;13:1398.
48. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88.
49. Smith AC, Thomas E, Snoswell CL, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). *J Telemed Telecare* 2020;26:309–313. <https://doi.org/10.1177/1357633X20916567>.
50. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. *J R Soc Med* 2019;112:22–28.
51. Feingold-Polak R, Elishay A, Shakir HS, et al. Robots' novel roles in physical and cognitive rehabilitation. *Front Psychol* 2019;10:1456.
52. Fong T, Nourbakhsh I, Dautenhahn K. A survey of socially interactive robots. *Robot Auton Syst* 2003;42:143–166.
53. Lebedev MA, Nicoletis MA. Brain-machine interfaces: from basic science to neuroprostheses and neurorehabilitation. *Physiol Rev* 2017;97:767–837.
54. Collinger JL, Wodlinger B, Downey JE, et al. High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* 2013;381:557–564. [https://doi.org/10.1016/S0140-6736\(12\)61816-9](https://doi.org/10.1016/S0140-6736(12)61816-9).
55. Savulescu J, Giubilini A, Vandersluis R, Mishra A. Ethics of artificial intelligence in medicine. *Singapore Med J* 2024;65:150–158.
56. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
57. Mesko B. *The role of artificial intelligence in precision medicine*. Taylor & Francis, 2017:239–241.
58. Lee S, Lieber L, Heck D. *Iron man Tales of suspense*. Marvel Comics, 1963.
59. Breedt J, Smits EJ, Haakma R, Twisk JW. Machine learning prediction of cognitive impairment in Parkinson's disease using neuropsychological tests and MRI data. *Front Aging Neurosci* 2022;14:851699.
60. Ziegler JC, Lallier M, Saadi LK, et al. Deep neural network modeling of visual word recognition and reading. *Neuropsychologia* 2022;171:108247.
61. Abedi A, Colella TJF, Pakosh M, Khan SS. Artificial intelligence-driven virtual rehabilitation for people living in the community: a scoping review. *npj Digit Med* 2024;7:25.
62. Wang X, Jiang L, Hernandez-Orallo J, et al. Evaluating general-purpose AI with psychometrics. *arXiv preprint arXiv:231016379*, 2023.
63. Bowers JS, Malhotra G, Dujmović M, et al. Deep problems with neural network models of human vision. *Behav Brain Sci* 2023;46:e385.
64. Shanahan M, Crosby M, Beyret B, Cheke L. Artificial intelligence and the common sense of animals. *Trends Cogn Sci* 2020;24:862–872.
65. Sueur C, Pelé M. Editorial: recent advances in animal cognition and ethology. *Animals* 2023;13:2890.
66. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018;1:18.
67. Avati A, Jung K, Harman S, et al. Improving palliative care with deep learning. *BMC Med Inform Decis Mak* 2018;18:55–64.
68. Weng SF, Reys J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One* 2017;12:e0174944.
69. Ren K, Qin J, Zheng L, et al. Deep recurrent survival analysis. *Proceedings of the AAAI conference on artificial intelligence*, 2019.
70. Clegg A, Young J, Iliffe S, et al. Frailty in elderly people. *Lancet* 2013;381:752–762.
71. Möckel T, Pletzer B, Huth J, et al. AI-based frailty assessment from clinical routine data predicts disability worsening and mortality in multiple sclerosis: an international multi-cohort observational study. *EBioMedicine* 2022;92:105918.
72. Pyrkov TV, Slipensky K, Barg M, et al. Extracting biological age from biomedical data via deep learning: too much of a good thing? *Sci Rep* 2018;8:5210.
73. Climent MT, Pardo J, Muñoz-Almaraz FJ, et al. Decision tree for early detection of cognitive impairment by community pharmacists. *Front Pharmacol* 2020;9:1232.
74. Stausholm MN, Egmoose A, Dahl SC, et al. Prediction of mortality in patients with heart failure using a novel machine learning algorithm. *Scand Cardiovasc J* 2019;56:338–343.
75. Heo J, Yoon JG, Park H, et al. Machine learning-based model for prediction of outcomes in acute stroke. *Stroke* 2019;50:1263–1265.
76. Rau C-S, Wu S-C, Chien P-C, et al. Prediction of mortality in patients with isolated traumatic subarachnoid hemorrhage using a decision tree classifier: a retrospective analysis based on a trauma registry system. *Int J Environ Res Public Health* 2017;14:1420.
77. Hernandez Rocha TA, Elahi C, Cristina da Silva N, et al. A traumatic brain injury prognostic model for human activity recognition using deep learning. *Sensors* 2021;17:616.
78. Royston P, Altman DG, Sauerbrei W. External validation and updating of a prognostic survival model. *Stat Med* 2015;34:2389–2400.
79. Zhang J, Song L, Chan KCG, et al. Machine learning models identify predictive features of patient mortality across dementia types. *Commun Med* 2024;4:23. <https://doi.org/10.1038/s43856-024-00437-7>.
80. De Pablo-Fernandez E, Tur C, Revesz T, et al. Association of autonomic dysfunction with disease progression and survival in Parkinson disease. *JAMA Neurol* 2017;74:970–976.
81. Agarwal S, Fleisher LA, Jain D. Machine learning to predict mortality in Parkinson's disease using gait and balance measures. *Sci Rep* 2021;11:11.
82. Ong M-L, Tan PF, Holbrook JD. Predicting functional decline and survival in amyotrophic lateral sclerosis. *PLoS One* 2017;12:e0174925.
83. Westeneng H-J, Debray TP, Visser AE, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol* 2018;17:423–433.
84. Tabrizi SJ, Ghosh R, Leavitt BR. Huntingtin lowering strategies for disease modification in Huntington's disease. *Neuron* 2019;101:801–819.
85. Illes J. *Neuroethics: anticipating the future*. Oxford University Press, 2017.
86. Coin A, Dubljević V. *Policy, identity, and neurotechnology: the Neuroethics of brain-computer interfaces*. Springer, 2023.

87. Farah MJ. *The neuroscience of ethics: A framework for ethical theory*. MIT Press, 2021.
88. Jotterand F, Ienca M. *Artificial intelligence in brain and mental health: philosophical, Ethical & Policy Issues*. Springer, 2021.
89. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2:230–243.
90. Glicksberg BS, Johnson KW, Tatonetti NP. Artificial intelligence and neurological diseases. *Lancet Neurol* 2018;17:1089–1090.
91. Panch T, Mattie H, Celi LA. The “inconvenient truth” about AI in healthcare. *npj Digit Med* 2019;2:77.
92. Yuste R, Goering S, Arcas BA, et al. Four ethical priorities for neurotechnologies and AI. *Nature* 2017;551:159–163.
93. Char DS, Shah NH, Magnus D. Implementing machine learning in health care-addressing ethical challenges. *N Engl J Med* 2018;378:981–983.
94. Thorne J, Vlachos A, Christodoulopoulos C, Mittal A. FEVER: a large-scale dataset for fact extraction and VERification, arXiv preprint arXiv:180305355. 2018.
95. Zellers R, Holtzman A, Rashkin H, et al. Defending against neural fake news. *Advances in neural information processing systems* 32, 2019.
96. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–453.
97. Hendrycks D, Burns C, Kadavath S, et al. Unsupervised translation of programming languages. arXiv preprint arXiv:210405634, 2021.
98. Hendrycks D, Zhao K, Basart S, et al. Natural adversarial examples. arXiv preprint arXiv:190707174, 2020.
99. Ridler C. Artificial intelligence accelerates detection of neurological illness. *Nature reviews. Neurology* 2018;14:572.
100. Qayyum SN, Ullah I, Rehan M, Noori S. AI integration in sepsis care: a step towards improved health and quality of life outcomes. *Ann Med Surg Lond* 2024;86:2411–2412.
101. Ross C, Swetlitz I. IBM's Watson recommended 'unsafe and incorrect' cancer treatments, internal reports show. *STAT*, 2018.
102. Stat. *UnitedHealth faced class action lawsuit over algorithmic care denials in Medicare advantage plans*. *STAT*, 2023.
103. Feuer WP. *AI hit with investor lawsuit over cancer diagnostic claims*. *CNBC*, 2022.
104. Donnelly L. Babylon Health faces legal case over missed brain aneurysm, 2021.
105. FDA. *FDA-approved artificial intelligence and machine learning (AI/ML)-enabled medical devices*. U.S. Food and Drug Administration, 2022.
106. Wu E, Wu K, Daneshjou R, et al. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med* 2021;27:582–584.
107. Dzindolet MT, Pierce LG, Beck HP, Dawe LA. Identifying inconsistency in human-machine interactions. *Proc Hum Factors Ergon Soc Ann Meet* 2003;47:441–445.
108. Maynez J, Narayan S, Patrini G, Christakopoulou K. Open-ended commonsense reasoning. arXiv preprint arXiv:200500733, 2020.
109. Novikova J, Dušek O, Curry AC, Rieser V. Why we need new evaluation metrics for NLG. arXiv preprint arXiv:170706875, 2018.
110. Talmor A, Herzig J, Lourie N, Berant J. CommonsenseQA: A question answering challenge targeting commonsense knowledge. arXiv preprint arXiv:181100937, 2020.
111. Burrell J. How the machine “thinks”: understanding opacity in machine learning algorithms. *Big Data Soc* 2016;3:2053951715622512.
112. Liopyris K, Gregoriou S, Dias J, Stratigos AJ. Artificial intelligence in dermatology: challenges and perspectives. *Dermatol Ther* 2022;12:2637–2651.
113. Institute of Medicine Committee on Quality of Health Care in A. In: Kohn LT, Corrigan JM, Donaldson MS, eds. *To err is human: building a safer health system*. Washington (DC): National Academies Press (US), 2000 Copyright 2000 by the National Academy of Sciences. All rights reserved.
114. Poore GD, Kopylova E, Zhu Q, et al. RETRACTED ARTICLE: microbiome analyses of blood and tissues suggest cancerdiagnostic approach. *Nature* 2020;579:567–574.
115. Arrieta AB, Díaz-Rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* 2020;58:82–115.
116. Turek M. *DARPA: Explainable Artificial Intelligence (XAI)*, 2020.
117. FDA. *Framework for the use of digital health technologies in drug and biological product development: how CBER, CDER, CDRH, and OCP are working together*. Guidance for industry, investigators, and other stakeholders, 2024.
118. Goldsack JC, Coravos A, Bakker JP, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). *NPJ Digit Med* 2020;3:55.
119. Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour. *Nature* 2019;568:477–486.
120. Xu K, Chen H, Nguyen KH, et al. V3: A system for validating, verifying, and visualizing robustness of AI systems. *AI Open* 2021;2:68–92.
121. Amodei D, Olah C, Steinhardt J, et al. Concrete problems in AI safety. arXiv preprint arXiv:160606565, 2016.
122. Dick PK. *Do androids dream of electric sheep?*. Doubleday, 1968.
123. Asimov I. *I Robot*. Gnome Press, 1950.
124. Robot I. [Film]. *Universal Pictures*. Directed by Alex Proyas, 2004.
125. Franklin S, Graesser A, eds. Is it an agent, or just a program?: A taxonomy for autonomous agents. *International workshop on agent theories, architectures, and languages*. Springer, 1997.
126. Searle JR. Minds, brains, and programs. *Behav Brain Sci* 1980;3:417–424.
127. Turing AM. *Computing machinery and intelligence*. Springer, 2009.
128. Heinlein RA. *Stranger in a Strange Land*. New York, New York: Penguin Publishing Group, 1961.
129. Power A, Burda Y, Edwards H, et al. Grokking: generalization beyond overfitting on small algorithmic datasets. arXiv preprint arXiv:220102177, 2022.
130. Boden MA. Creativity and artificial intelligence. *Artif Intell* 1998;103:347–356.
131. Minsky M. *The emotion machine*. Simon and Schuster, 2006.
132. Damasio AR. *The feeling of what happens: body and emotion in the making of consciousness*. Houghton Mifflin Harcourt, 1999.
133. Harnad S. The Turing test is not a trick: Turing indistinguishability is a scientific criterion. *ACM SIGART Bull* 1992;3:9–10.
134. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Netw* 2015;61:85–117.
135. Baron-Cohen S, Leslie A, Frith U. Does the autistic child have a “theory of mind”? *Cognition* 1985;21.
136. Wallach W, Allen C. *Moral machines: teaching robots right from wrong*. Oxford University Press, 2008.
137. Dehaene S. *Consciousness and the brain: deciphering how the brain codes our thoughts*. Penguin, 2014.
138. Tononi G. The integrated information theory of consciousness: an updated account. *Arch Ital Biol* 2012;150:56–90.

139. Baars BJ. *A cognitive theory of consciousness*. Bernard Baars, 1988.
140. Bryson JJ. Robots should be slaves. *Close engagements with artificial companions: Key social, psychological, ethical and design issues*. Vol 8. Amsterdam: John Benjamins Publishing Company, 2010:63-74.
141. Čapek K. *R.U.R. (Rossum's universal robots)*. Prague: Aventinum, 1920.
142. Kurzweil R. The singularity is near. In: Sandler RL, ed. *Ethics and emerging technologies*. London: Palgrave Macmillan UK, 2014: 393-406.
143. Bostrom N. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.
144. Floridi L, Cowls J, Beltrametti M, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds Mach* 2018;28:689-707.
145. International S. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, 2021. Available at: https://www.sae.org/standards/content/j3016_202104/.
146. Yanco HA, Drury JL. Classifying human-robot interaction: an updated taxonomy, Paper presented at: Proceedings of the IEEE International Conference on Systems, Man & Cybernetics: The Hague, Netherlands, 10-13 October 2004.
147. Fong T, Thorpe C, Baur C, eds. Collaboration, dialogue, human-robot interaction. *Robotics research: The tenth international symposium*. Springer, 2003.
148. Goodrich MA, Schultz AC. Human-robot interaction: a survey. *Foundations and trends® in human-computer. Interaction* 2008;1: 203-275.
149. Beer JM, Fisk AD, Rogers WA. Toward a framework for levels of robot autonomy in human-robot interaction. *J Hum Robot Interact* 2014;3:74-99.
150. Klien G, Woods DD, Bradshaw JM, et al. Ten challenges for making automation a “team player” in joint human-agent activity. *IEEE Intell Syst* 2004;19:91-95.
151. Scholtz J. Theory and evaluation of human robot interactions. *36th Annual Hawaii International Conference on System Sciences, 2003 Proceedings of the. IEEE*, 2003.
152. Institute PM. *A guide to the project management body of knowledge (PMBOK guide)*. 7th ed. Project Management Institute, 2021.
153. Winkle K, Guillory S. Social robots and human interaction: A systematic review of current research. *Int J Soc Robot* 2023;15:213-230.
154. Adamson H, Zlotowski J. Designing emotionally intelligent AI: A review of approaches and applications. *J Hum Robot Interact* 2023; 12:45-68.
155. Bao Y, Gong W, Yang K. A literature review of human-AI synergy in decision making: from the perspective of affordance actualization.
156. No Robot Bosses Act, 2024.
157. Steinmetz JD, Seeher KM, Schiess N, et al. Global, regional, and national burden of disorders affecting the nervous system, 1990-2021: a systematic analysis for the global burden of disease study 2021. *Lancet Neurol* 2024;23:344-381.
158. Sakowski SA, Koubek EJ, Chen KS, et al. Role of the exposome in neurodegenerative disease: recent insights and future directions. *Ann Neurol* 2024;95:635-652.
159. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;2:3. <https://doi.org/10.1186/2047-2501-2-3>.
160. Frankovich J, Longhurst CA, Sutherland SM. Evidence-based medicine in the EMR era. *N Engl J Med* 2011;365:1758-1759.
161. WHO. *International statistical classification of diseases and related health problems, 10th Revision (ICD-10)*. Geneva: WHO: World Health Organization, 2022.
162. International. S. *SNOMED clinical terms (SNOMED CT)*. London: SNOMED International, 2023.
163. Medicine UNLo. *RxNorm*. Bethesda: NLM: U.S. National Library of Medicine, 2023.
164. Institute. R. *LOINC-logical observation identifiers names and codes*. Indianapolis: Regenstrief Institute, 2023.
165. Bender D, Sartipi K. HL7 FHIR: an agile and RESTful approach to healthcare information exchange. *Proceedings of the 26th IEEE international symposium on computer-based medical systems*. IEEE, 2013.
166. Cella D, Gershon R, Bass M, Rothrock N. Assessment center scoring service: A resource for those seeking to promote physical, mental and social health. *J Appl Meas* 2019;20:61-73.
167. Observational Health Data Sciences and Informatics (OHDSI). (n.d.). *Standardized Data: The OMOP Common Data Model*. Retrieved May 3, 2025, from <https://www.ohdsi.org/data-standardization/>.
168. Neville J, Khlif MS, Bemedrisov A, et al. FHIR server and data transformation for streamlining ADNI data sharing and analysis. *J Alzheimers Dis* 2021;79:703-713.
169. Lyons J, Akhauri S, Khrac N, et al. Massively parallel phenotyping of Alzheimer's disease risk genes. *Neuron* 2021;109:2704-2720.
170. Petters D, Thun S, Serna E, Gomez-Perez JM. Ontology engineering methodology for neurocognitive disorders. *J Biomed Semant* 2019;10:9.
171. Bilder RM, Sabb F, Cannon T, et al. Phenomics: the systematic study of phenotypes on a genome-wide scale. *Neuroscience* 2009; 164:30-42.
172. Dutta B. Theoretical analysis and propositions for “ontology citation”, 2018. Paper presented at: Proceedings of the international conference on exploring the horizons of library and information sciences: from libraries to knowledge hubs, 7-9 August, 2018 Bangalore, India.
173. Price CJ, Friston KJ. Functional ontologies for cognition: the systematic definition of structure and function. *Cogn Neuropsychol* 2005;22:262-275.
174. Poldrack RA, Kittur A, Kalar D, et al. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Front Neuroinform* 2011;5:17. <https://doi.org/10.3389/fninf.2011.00017>.
175. Sargsyan A, Wegner P, Gebel S, et al. The epilepsy ontology: a community-based ontology tailored for semantic interoperability and text mining. *Bioinform Adv* 2023;3:vb033.
176. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform* 2019;95:103208.
177. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377-381.
178. Yang X, Chen A, Pourmejian N, et al. GatorTronGPT: a generative large language model for medical research. *Nat Commun* 2023;14: 5678. <https://doi.org/10.1038/s41746-023-00958-w>.
179. Papamakarios G, Nalisnick E, Rezende DJ, et al. Normalizing flows for probabilistic modeling and inference. *J Mach Learn Res* 2021; 22:1-64.
180. Beizae F, Desrosiers C, Lodygensky GA, Dolz J. Leveraging normalizing flows for unsupervised and source-free MRI harmonization. *arXiv preprint arXiv:2407.15717*, 2024. <https://doi.org/10.48550/arXiv.2407.15717>.
181. Čapek K. R.U.R (Rossum's Universal Robots) Poster [Poster], 1920.
182. Čapek K. R.U.R (Rossum's Universal Robots) Scene 1 [Photo], 1920.
183. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Adv Neural Inf Process Syst* 2014;27:2672-2680.

184. Ahmad B, Usama M, Hussain M, et al. Brain tumor classification using a combination of variational autoencoders and generative adversarial networks. *Biomedicines* 2022;10:223. <https://doi.org/10.3390/biomedicines10020223>.
185. Chen H, Dou Q, Yu L, et al. CycleGAN-based PET synthesis for improved Alzheimer's disease diagnosis. *IEEE Trans Med Imaging* 2020;39:2549–2559.
186. Li Y, Yakushev I, Hedderich DM, Wachinger C. PASTA: Pathology-Aware MRI to PET Cross-Modal Translation with Diffusion Models. arXiv preprint arXiv:2405.16942, 2024.
187. Scarfone K, Hoffman P. Guidelines on firewalls and firewall policy (NIST Special Publication 800-41 Rev. 1). *National Institute of Standards and Technology*, 2009. <https://doi.org/10.6028/NIST.SP.800-41r1>.